I. AMENDMENT

Amendment of the Specification

The paragraph beginning on page 8 line 1 is re-written as follows:

Fig. 5: Amino acid sequence comparison of the binding protein and the inhibitor of human hepatocyte growth factor activator (HAI-1). Twelve-amino acid (GPPPAPPGLPAG) (SEQ ID NO: 2) and seven-amino acid (TQGFGGS) (SEQ ID NO: 3) sequences of the amino termini obtained from the 40-kDa binding protein doublet and the 25-kDa binding protein, respectively, and were identical to amino acids 36-47 and 154-160 of HAI-1 (SEQ ID NO: 1). In addition, 12 unique peptides from the tryptic digest of the larger band of the 40-kDa binding protein doublet were compared with HAI-1 by MALDI-MS. These peptides covered 87 residues that spanned positions 135-310, or 17% of the entire HAI-1. The two stretches of amino-terminal protein sequences are double-underlined, and those 12 peptides identified by MALDI-MS, including residues 135-143, 154-164, 165-172, 173-182, 173-190, 183-190, 194-199, 203-214, 204-214, 288-301, and 302-310 (SEQ ID NO: 1), are underlined.

The paragraph beginning on page 10 line 10 is re-written as follows:

Fig. 9: The nucleotide and deduced amino acid sequences (SEQ ID NO: [[3]] 4) of a matriptase cDNA clone. The primers (20 bases at the 5'-end and 18 bases at the 3'- end) used for reverse transcriptase-polymerase chain reaction are underlined. Thirty-three bases beyond the 5'-end primer and 92 bases beyond the 3'-end primer were taken from SNC19 cDNA and incorporated. The cDNA sequence (SEQ ID NO: [[1]] 5) was translated from the fifth ATG codon in the open reading frame. Nucleotide and amino acid numbers are shown on the left. Sequences that agreed with the internal sequences obtained from matriptase are double-underlined. His-484, Asp-539, and Ser-633 are boxed and indicated the putative catalytic triad of matriptase. Potential N-glycosylation sites ~~an~~ are indicated (Δ). An RGD sequence is indicated (♠).

The paragraph beginning on page 10 line 21 is re-written as follows:

Fig. 10: Comparison of the amino acid sequence of the C-terminal region of matriptase with trypsin, chymotrypsin, and with the catalytic domains of other serine proteases. The C-terminal region (amino acids 431-683) (SEQ ID NO: 6) of matriptase is compared with human trypsin (SEQ ID NO: 13) (Emi *et al., Gene (Amst.)* 41: 305-10 (1986)); human chymotrypsin (SEQ ID NO: 14) (Tomita *et al., Biochem. Biophys. Res. Commun.* 158: 569-75 (1989)); the catalytic chains of human enteropeptidase (SEQ ID NO: 7) (Kitamoto *et al., Proc. Natl. Acad. Sci. USA* 91: 7588-92 (1994)), human hepsin (SEQ ID NO: 10) (Leytus *et al., Biochemistry* 27: 1067-74 (1988)), human blood coagulation factor XI (SEQ ID NO: 11) (Fujikawa *et al., Biochemistry* 25: 2417-24 (1986)), and human plasminogen (SEQ ID NO: 12); and the serine protease domains of two transmembrane serine proteases, human TMPRSS2 (SEQ ID NO: 8) (Paoloni-Giacobino *et al., Genomics* 44: 309-20 (1997)) and the *Drosophila Stubble-stubbloid* gene (*Sb-sbd*) (SEQ ID NO: 9) (Appel *et al., Proc. Natl. Acad. Sci. USA* 90: 4937-41 (1993)). Gaps to maximize homologies are indicated by dashes. Residues in the catalytic triads (matriptase His-484., Asp-539, and Ser-633) are boxed and indicated (▲). The conserved activation motif ((R/K)VIGG) (SEQ ID NO: 34) is boxed, and the proteolytic activation site is indicated. Eight conserved cysteines needed to form four intramolecular disulfide bonds are *boxed*, and the likely pairings are as follows: Cys-469-Cys-485, Cys-604-Cys-618, Cys-629-Cys-658, and Cys-432-Cys-559. The disulfide bond Cys-432-Cys-559. The disulfide bond Cys-432-Cys-559 is observed in two-chain serine proteases, but not in trypsin and chymotrypsin. Residues in the substrate pocket (Asp-627, Gly-655, and Gly-665) are *boxed* and indicated (♣). It is evident that the residue positioned at the bottom of the substrate pocket is Asp in trypsin-like proteases, including matriptase, but Ser in chymotrypsin.

The paragraph beginning on page 11 line 19 is re-written as follows:

Fig. 11: Alignment of partial sequences of the noncatalytic domain with those of homologous regions in other proteins. *A*, the cysteine-rich repeats of matriptase (amino acids 280-314, 315-351, 352-387, and 394-430) (SEQ ID NO: 15) are compared with the consensus sequences of the human LDL receptor (SEQ ID NO: 16) (Sudhof *et al., Science*

228: 815-22 (1985)), LDL receptor-related protein (*LRP*) (SEQ ID NO: 17) (Herz *et al.*, *EMBO J.* 7: 4119-27 (1988)), human perlecan (SEQ ID NO: 18) (Murdoch *et al.*, *J. Biol. Chem.* 267: 8544-57 (1992)), and rat GP-300 (SEQ ID NO: 19) (Raychowdhury *et al.*, *Science* 244: 1163-65 (1989)). The consensus sequences are boxed. *B*, C1r/s-type sequences of matriptase (*Mt*; amino acids 42-155 and 168-268) (SEQ ID NOS: 20 and 21) are compared with selected domains of human complement subcomponent C1r (amino acids 193-298) (SEQ ID NO: 22) (Leytus *et al.*, *Biochemistry* 25: 4855-63 (1986); Journet, *Biochem. J.* 240: 783-87 (1986)), C1s (amino acids 175-283) (SEQ ID NO: 23) (Mackinnon *et al.*, *Eur. J. Biochem.* 169: 547-53 (1987); and Tosi *et al.*, *Biochemistry* 26: 8516-24 (1987)), Ra-reactive factor (*RaRF*) (amino acids 185-290) (SEQ ID NO: 24) (Takada *et al.*, *Biochem. Biophys. Res. Commun.* 196: 1003-9 (1993); and Sato *et al.*, *Int. Immunol.* 6: 665-9 (1994)), and a calcium dependent serine protease (*CSP*) (amino acids 181-289) (SEQ ID NO: 25) (Kinoshita *et al.*, *FEBS Lett.* 250: 411-5 (1989)). The consensus sequences are boxed.

The paragraph beginning on page 13 line 15 is re-written as follows:

Fig. 15: Nucleic acid sequence for human matriptase (SEQ ID NO: [[2]]1 26). SEQ ID NO: [[2]] 26 contains additional nucleic acid sequence encoding the first 172 amino acids located in the amino-terminus of the encoded protein as compared to SEQ ID NO: [[1]] 4, which [[is]] encodes a truncated form of matriptase. SEQ ID NO: [[2]] 26 represents the full-length form of the nucleic acid encoding matriptase, whereas SEQ ID NO: [[1]] 4 [[is]] encodes a truncated form. The sequence can be found at GenBank Accession No. AF118224.

The paragraph beginning on page 13 line 22 is re-written as follows:

Fig. 16: Amino acid sequence for human matriptase (SEQ ID NO: [[4]] 27). This sequence contains 855 amino acids, which is larger than the sequence described in Lin *et al.*, J. Biol. Chem. 274: 18231-6 (1999) (SEQ ID NO: [[2]]5). SEQ ID NO: [[4]] 27 is the full length form of the matriptase protein, whereas SEQ ID NO: [[3]] 5 is a truncated matriptase protein lacking 172 amino acids at the amino terminus. The protein sequence can be found at GenBank Accession No. AAD42765.

The paragraph beginning on page 14 line 17 is re-written as follows:

Another object of the invention is to provide nucleic acid molecules (SEQ ID NOS: [[1]] 4 and [[2]] 26) encoding matriptase proteins or polypeptide fragments thereof (SEQ ID NOS: [[3]] 5 and [[4]] 27).

The paragraph beginning on page 14 line 20 is re-written as follows:

It is a further object of the invention to provide an antibody or antibodies which recognizes and binds to SEQ ID NO: [[3]] 5 or a fragment thereof, SEQ ID NO: [[4]] 27 or a fragment thereof, to a single-chain (zymogen) form of matriptase or to a two-chain (active) from of matriptase. Preferred antibodies are monoclonal antibodies and fragments thereof as well as chimeric, humanized or human antibodies.

The paragraph beginning on page 17 line 21 is re-written as follows:

By "matriptase" is meant a trypsin-like protein, with a molecular weight of between 72-kDa and 92-kDa and is related to SEQ ID NO: [[4]] 27 or is a fragment thereof. It can include both single-chain and double-chain forms of the protein. The zymogen form (inactive form) of matriptase is a single-chain protein. The two-chain form of matriptase is the active form of matriptase, which possesses catalytic activity. Both forms of matriptase are found to some extent in milk and cancer cells because extracellular matrix (ECM) remodeling is necessary to both normal and pathologic remodeling processes. Figure 14 displays all known forms of matriptase. Both cancer cells and milk contain the different forms of matriptase. However, in milk the dominant form is the activated form of matriptase which then binds to HAI-1.

The paragraph beginning on page 22 line 9 is re-written as follows:

The present invention further provides nucleic acid molecules that encode the protein having SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27, or fragments thereof, and related proteins, which are preferably in isolated form. As used herein, "nucleic acid" is defined as RNA or DNA that encodes a peptide as defined above, or is complementary to nucleic acid sequence encoding such peptides, or hybridizes to such nucleic acid and remains stably

bound to it under appropriate stringency conditions, or encodes a polypeptide sharing at least 75% sequence identity, preferably at least 80%, and more preferably at least 85%, with the peptide sequences. Specifically contemplated are genomic DNA, cDNA, mRNA and antisense molecules, as well as nucleic acids based on alternative backbone or including alternative bases whether derived from natural sources or synthesized.

The paragraph beginning on page 24 line 14 is re-written as follows:

As described above, the identification of the human nucleic acid molecule having SEQ ID NO: [[1]] 4 or SEQ ID NO: [[2]] 26 allows a skilled artisan to isolate nucleic acid molecules that encode other members of the matriptase family, in addition to the human sequence herein described. Further, the presently disclosed nucleic acid molecules allow a skilled artisan to isolate nucleic acid molecules that encode other members of the matriptase family of proteins in addition to the disclosed protein having SEQ ID NO: [[3]] 5 and SEQ ID NO: [[4]] 27.

The paragraph beginning on page 24 line 21 is re-written as follows:

Essentially, a skilled artisan can readily use the amino acid sequence of NO: [[3]] 5 or SEQ ID NO: [[4]] 27 to generate antibody probes to screen expression libraries prepared from appropriate cells. Typically, polyclonal antiserum from mammals, such as rabbits, immunized with the purified protein (as described below) or monoclonal antibodies can be used to probe a mammalian cDNA or genomic expression library, such as *λgtll* library, to obtain the appropriate coding sequence for other members of the protein family. The cloned cDNA sequence can be expressed as a fusion protein, expressed directly using its own control sequences, or expressed by constructions using control sequences appropriate to the particular host used for the expression of the enzyme.

The paragraph beginning on page 29 line 9 is re-written as follows:

First, a nucleic acid molecule is obtained that encodes a protein of the invention, such as the nucleic acid molecule depicted in SEQ ID NOS: [[1]] 4 or [[2]] 26, or particularly for the matriptase nucleotides encoding for example, the serine protease catalytic domain in the carboxy terminus of the matriptase protein or the LDL domain. The coding sequence is

directly suitable for expression in any host, as it is not interrupted by introns. The sequence can be transfected into host cells such as eukaryotic cells or prokaryotic cells. Eukaryotic hosts include mammalian cells (*e.g.*, HEK293 cells, CHO cells and PAE-PDGF-R cells), as well as insect cells such as Sf9 cells using recombinant baculovirus. Alternatively, fragments encoding only portion of matriptase can be expressed alone or as a fusion protein. For example, the C-terminus of matriptase containing the serine protease domain can be expressed in bacteria as a GST- or His-tag fusion protein. These fusion proteins can then purified and used to generate polyclonal antibodies.

The paragraph beginning on page 30 line 16 is re-written as follows:

Another embodiment of the present invention provides methods for use in isolating and identifying binding partners of matriptase proteins. In detail, a protein of the invention is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the protein of the invention. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a protein of the invention are separated from the mixture. The binding partner that bound to the protein of the invention can then be removed and further analyzed. To identify and isolate a binding partner, the entire protein, for instance the entire disclosed protein of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27 can be used. Alternatively, a fragment of the protein can be used.

The paragraph beginning on page 33 line 3 is re-written as follows:

Another embodiment of the present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding a protein of the invention, such as a protein having the amino acid sequence of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27. Such assays may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a nucleic acid of the invention, for instance a nucleic acid encoding the protein having the sequence of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27, if it is capable of up- or down-regulating expression of the nucleic acid in a cell.

The paragraph beginning on page 33 line 12 is re-written as follows:

In one assay format, cell lines that contain reporter gene fusions between the open reading frame of matriptase or of SEQ ID NOS: [[1]] 4 or [[2]] 26 and any assayable fusion partner may be prepared. Numerous assayable fusion partners are known and readily available including the firefly luciferase gene and the gene encoding chloramphenicol acetyltransferase (Alam *et al., Anal. Biochem.* 188: 245-54 (1990)). Cell lines containing the reporter gene fusions are then exposed to the agent to be tested under appropriate conditions and time. Differential expression of the reporter gene between samples exposed to the agent and control samples identifies agents which modulate the expression of a nucleic acid encoding the protein having the sequence of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27 or related proteins.

The paragraph beginning on page 33 line 23 is re-written as follows:

Additional assay formats may be used to monitor the ability of the agent to modulate the expression of a nucleic acid encoding a protein of the invention such as the protein having SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27. For instance, mRNA expression may be monitored directly by hybridization to the nucleic acids of the invention. Cell lines are exposed to the agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such those disclosed in Sambrook *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed. Cold Spring Harbor Laboratory Press, 1989). Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared from the nucleic acids of the invention. It is preferable, but not necessary, to design probes which hybridize only with target nucleic acids under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the amount of complementarity which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the difference in stability between the probe:target hybrid and potential probe:non-target hybrids.

The paragraph beginning on page 34 line 20 is re-written as follows:

Hybridization conditions are modified using known methods, such as those described by Sambrook *et al.* (1989) and Ausubel *et al.* (1995), as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA RNA can be accomplished in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support, and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively, nucleic acid fragments comprising at least one, or part of one of the sequences of the invention can be affixed to a solid support, such as a porous glass wafer. The glass wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize. Such glass wafers and hybridization methods are widely available, for example, those disclosed by Beattie (WO 95/11755). By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up or down regulate the expression of a nucleic acid encoding the protein having the sequence of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27 are identified.

The paragraph beginning on page 35 line 13 is re-written as follows:

Another embodiment of the present invention provides methods for identifying agents that modulate at least one activity of a protein of the invention, such as the protein having the amino acid sequence of SEQ ID NO: [[3]] 5 or SEQ ID NO: [[4]] 27. Such methods or assays may utilize any means of monitoring or detecting the desired activity.

The paragraph beginning on page 37 line 110 is re-written as follows:

While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, use of monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using the standard method of Kohler *et al.*, (*Nature* 256: 495-7 (1975)) or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The

immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production *in vivo* via ascites fluid. Of particular interest, are monoclonal antibodies which recognize the catalytic domain of matriptase (*e.g.*, amino acids 432-683 of SEQ ID NO: [[3]] 5).

The paragraph beginning on page 53 line 3 is re-written as follows:

Expression of HAI-1 in COS-7 cell: To verify that HAI-1 encodes the binding protein of matriptase, we isolated an HAI-1 cDNA fragment by reverse transcriptase-polymerase chain reaction (RT-PCR) utilizing mRNA from MTSV 1.1B immortalized human luminal mammary epithelial cells. Primer sequences for HAI-1 (5'-GGCCCGCGCTCTGAAGGTGA -3' (SEQ ID NO: 28) and 5'-TTGGCAAGCAGGAAGCAGGG-3') (SEQ ID NO: 29) were derived from the published sequence. Standard RT-PCR was carried out using the Advantage RT-PCR kit (Clontech; Palo Alto, CA), and the product was ligated into pCR2.1 (Invitrogen; Carlsbad, CA) by TA cloning. The sequence of the RT-PCR product was obtained by standard methods, and confirmed with the published HAI-1 sequence (Miyazawa *et al., J. Biol. Chem.* 268: 10024-8 (1993)). An eukaryotic expression vector was constructed (pcDNA/HAI-1), utilizing the commercially available pcDNA3.1 vector (Invitrogen; San Diego, CA). A 1.6 kb EcoRI fragment containing the HAI-1 cDNA was cloned into the EcoRI site of pcDNA 3.1. This construct contains the open reading frame (ORF) of HAI-1 driven by a CMV promoter. Correct insertion of the HAI-1 cDNA was verified by restriction mapping. Transfections were performed using SuperFect transfection reagent (QIAGEN; Valencia, CA) as specified in manufacturer's handbook. After 48 hr, the HAI-1-transfected COS-7 cells were extracted with 1% Triton-X100 in 20 mM Tris-HCl pH 7.4.

The paragraph beginning on page 59 line 19 is re-written as follows:

Amplification of an SNC19 CDNA from T-47D breast cancer cells: An SNC19 cDNA clone was generated by reverse transcriptase-polymerase chain reaction (RT-PCR) utilizing mRNA from T-47D human breast cancer cells. Primer sequences for SNC19 (5'-CCTCCTCTTGGTCTTGCTGGGG-3' (SEQ ID NO: 30) and 5'-AGACCCGTCTGTTTT

CCAGG-3') (SEQ ID NO: 31) were derived from the published sequence. Standard RT-PCR was conducted using the Advantage RT-PCR kit (Clontech; Palo Alto, CA). Products were analyzed on a 0.8% agarose gel and the resultant band of approximately 2.8 kb corresponding to the expected product size was excised from the gel, purified and ligated into pCR2.1 (Invitrogen, Carlsbad, CA) by TA cloning (pCR-SNC19).

The paragraph beginning on page 62 line 22 is re-written as follows:

Although sequence analysis of the 40-kDa binding protein has shown it to be a serine protease inhibitor (see below), some residual gelatinolytic activity was observed for the 95-kDa matriptase/inhibitor complex (Fig. 8 C). When matriptase and its binding protein were subjected to N-terminal sequencing, only 11 amino acid residues (VVGGTDADEGE) (SEQ ID NO: 32) from matriptase were obtained with relatively low recovery, and 12 amino acid residues (GPPPAPPGLPAG) (SEQ ID NO: 2) were obtained from the amino-terminus of the 40-kDa binding protein ~~have been obtained~~. The 11 amino acid residues from matriptase were identical to a deduced amino acid sequence from a 2.9 kb cDNA called SNC19 (accession number U20428). Numerous stop codons were observed in this deposited SNC19 sequence, resulting in several small, predicted translation products. Thus, a 2,830 bp cDNA fragment was obtained by reverse transcriptase-polymerase chain reaction using two primers based on the sequence of SNC 19. There was extensive discrepancy (132 bases) between our sequence and that of SNC19.

The paragraph beginning on page 63 line 19 is re-written as follows:

Nucleotide and predicted amino acid sequences of an matriptase cDNA clone: A nucleotide (SEQ ID NO: [[1]] 4) and an amino acid ~~sequences~~ sequence (SEQ ID NO: [[3]] 5) of matriptase are shown in Fig. 9. The primers (20 bases at 5' end and 18 bases at 3' end) used for reverse transcriptase-polymerase chain reaction are underlined. Thirty three bases beyond the 51 end primer and 92 bases beyond 31 end primer were taken from SNC 19 cDNA and incorporated. The cDNA sequence was translated from the fifth ATG (Met) codon in the open reading frame. Nucleotide and amino acid numbers are shown on the left. Double-underlines indicate sequences that agreed with the internal sequences obtained from matriptase. His-484, Asp-539 and Ser-633 were boxed and indicated the putative catalytic

triad of matriptase. Potential N-glycosylation sites are indicated by Δ. A RGD sequence is indicated by ♤.

The paragraph beginning on page 64 line 21 is re-written as follows:

Matriptase cDNA is likely to be 2955 base pair long when the 5' end 33 bases and the 3' end 92 bases from SNC 19 were added to the RT-PCR fragment (2,830 base pair long). The translation initiation site was assigned to the fifth methionine codon because the sequence GTCATGG matches a favorable Kozak consensus sequence (Kozak *et al.*, Nucl. Acid. Res. 12: 857-72 (1984)). This methionine is followed by four positively charged amino acids and a 14 amino acid long hydrophobic region (Ser-18-Ser-31), a putative signal peptide. Assuming this methionine codon to be the initiator, the open reading frame was 2,049 base pairs long, and thus the deduced amino acid sequence was composed of 683 residues, with calculated molecular mass of 75,626. The two stretches of amino acid sequences (DYVEINGEK (SEQ ID NO: 33) and VVGGTDADEGE) (SEQ ID NO: 32) obtained from matriptase are located in aa 228-236 and aa 443-453; thus the translation frame is likely to be correct. There are three potential N-glycosylation sites with the canonical Asn-X-(Ser/Thr) and an RGD sequence. RGD sequence from proteins of the extracellular matrix has been found to mediate interactions with integrins (Ruoslahti *et al.*, Science 238: 491-7 (1987)).

The paragraph beginning on page 64 line 21 is re-written as follows:

Structure of the matriptase catalytic domain: A homology search for the deduced amino acid sequence by BLAST in the Swiss-Prot© data base reveals that (1) the carboxyl-terminus at residue positions 432-683 of matriptase (SEQ ID NO: 5) is homologous with other serine proteases; (2) matriptase contains the invariant catalytic triad; (3) matriptase contains a characteristic disulfide bond pattern; and (4) matriptase contains overall sequence similarity. Referring to Figure 9, the primers (20 bases at 5' end (SEQ ID NO: 30) and 18 bases at 3' end (SEQ ID NO: 31, complement)) used for reverse transcriptase-polymerase chain reaction are underlined. Thirty-three bases beyond the 5' end primer and 92 bases beyond 3' end primer were taken from SNC19 cDNA and incorporated. The cDNA sequence was translated from the fifth ATG codon in the open reading frame. Nucleotide and amino acid numbers are shown on the left. Double-underlines indicate sequences that agreed with

the internal sequences obtained from matriptase. His-484, Asp-539, and Ser-633 were boxed and indicated the putative catalytic triad of matriptase. Potential N-glycosylation sites are indicated by Δ. A RGD sequence is indicated by ♤.

The paragraph beginning on page 66 line 3 is re-written as follows:

Referring more specifically to Figure 10, the C-terminal region (aa 431-683) of matriptase (SEQ ID NO: 6) is compared with human trypsin (SEQ ID NO: 13), human chymotrypsin (SEQ ID NO: 14), the catalytic chains of human enteropeptidase (SEQ ID NO: 7), human hepsin (SEQ ID NO: 10), human blood coagulation factor XI (SEQ ID NO: 11), and human plasminogen (SEQ ID NO: 12), and the serine protease domains of two transmembrane serine proteases, human TMPRSS2 (SEQ ID NO: 8) and Drosophila Stubble-stubbloid gene (Sb-sbd) (SEQ ID NO: 9). Residues are expressed in one letter code. Gaps to maximize homologies are indicated by residues in the catalytic triads (matriptase His-484, Asp-539, and Ser-633) were boxed and indicated by ♦. The conserved activation motif (R/KVIGG) (SEQ ID NO: 34) was boxed and the proteolytic activation site was indicated. Eight conserved cysteines needed to form four intramolecular disulfide bonds are boxed, and the likely pairings are as follows: Cys-469-Cys-485, Cys-604-Cys-618, Cys-629-Cys-658, and Cys-432-Cys-559. The disulfide bond (Cys-432-Cys-559) is observed in two-chain serine proteases, but not in trypsin and chymotrypsin. Residues in the substrate pocket (Asp-627, Gly-655, and Gly-665) are boxed and indicated by ♧. It is evident that the residue positioned at the bottom of substrate pocket is Asp in trypsin-like proteases, including matriptase, but is Ser in chymotrypsin.

The paragraph beginning on page 66 line 20 is re-written as follows:

A putative proteolytic activation site (Arg-442) of matriptase in a motif of Arg-Val-Val-Gly-Gly (RVVGG) (SEQ ID NO: 35) is similar to the characteristic RIVGG (SEQ ID NO: 36) motif in other serine proteases. However, the Ile residue is replaced by Val residue. This replacement is uncommon, but is observed in plasminogen. As mentioned above, a conserved intramolecular disulfide bond is found in those serine proteases that are synthesized as one-chain zymogens and are proteolytically activated to become active two chain forms. This disulfide bond is proposed to hold together the active catalytic fragment

with their non-catalytic N-terminal fragments, thus serving as protein-protein interaction domain. This conserved intramolecular disulfide bond has been also observed in matriptase (Cys-432-Cys-559). These sequence analyses suggest that matriptase may be synthesized as a single chain zymogen and may become proteolytically activated to a two-chain form. If this is a case, the majority of matriptase in the conditioned medium of T-47D breast cancer cells is likely to be the zymogen; the active two-chain matriptase only represents a minor proportion, consistent with the purified matriptase from T-47D human breast cancer cells exhibiting an apparent size of 80-kDa under reduced conditions. This conclusion is also supported by the observation that the proposed N-terminal sequences for the catalytic chain of matriptase are identical to the stretch of amino acid sequences (VVGGTDADEGE) (SEQ ID NO: 37) , which were obtained with very low recovery when matriptase was subjected to N-terminal sequencing.

The paragraph beginning on page 67 line 14 is re-written as follows:

The substrate specificity (S_1) pocket of matriptase is likely to be composed of Asp-627 positioned at its bottom, with Gly-655 and Gly-665 at its neck, indicating that matriptase is a typical trypsin-like serine protease. Predicted preferential cleavage at amino acid residues with positively charged side chains was confirmed with various synthetic substrates with Arg and Lys residues as P1 sites [[(Fig. 11)]]. Specifically, matriptase was able to cleave the following synthetic substrates, presented as follows, from the most rapid to the slowest: Boc-Gln-AlaArg-AMC, Boc-benzyl-Glu-Gly-Arg-AMC, Boc-Leu-Gly-Arg-AMC, Bocbenzyl-Asp-Pro-Arg-AMC, Boc-Phe-Ser-Arg-AMC, Boc-Leu-Arg-Arg-AMC, Boc-Gly-Lys-Arg-AMC, and Boc-Leu-Ser-Thr-Arg-AMC. [Boc = *t*-butyloxycarbonyl; AMC = 7-amino-4-methylcoumarin; LDL = low density lipoprotein]. This supports the view that matriptase prefers substrates with amino acid residues containing small side chains, such as Ala and Gly as P2 sites. These results suggest that matriptase, in analogy with trypsin. exhibits broad spectrum cleavage specificity. This broad spectrum cleavage activity is likely to be the explanation of its gelatinolytic activity. Its trypsin-like activity appears to be distinct from Gelatinases A and B, which may cleave gelatin at glycine residues, the most abundant (almost up to one third of) amino acid residues in gelatin.

The paragraph beginning on page 68 line 6 is re-written as follows:

Structure motifs of the noncatalytic region of matriptase: The non-catalytic region of matriptase contains two sets of repeating sequences, which may serve as a regulatory and/or binding domain for interaction with other proteins. Four tandem repeats of about 35 amino acids including 6 conserved cysteine residues (Fig. [[12 A]] 11 A) were found at the amino terminal region (aa 280-430) of its serine protease domain. They are homologous with the cysteine-containing repeat of the LDL receptor (Sudhof *et al.*, Science 228: 815-22 (1985)) and related proteins (Herz *et al.*, EMBO J. 7: 4119-27 (1988)). All of these cysteine residues are likely be involved in disulfide bonds. In LDL receptor, the homologous, seven repeating sequences serve as the ligand binding domain. By analogy, the four tandem cysteine-containing repeats may also be the sites of interaction with other macromolecules. In addition, the cysteine-containing LDL receptor domain was found in other proteases, such as enteropeptidase (Matsushima *et al.*, J. Biol. Chem. 269: 19976-82 (1994); and Kitamoto *et al.*, Proc. Nad. Acad. Sci. USA 91: 7588-92 (1994)).

The paragraph beginning on page 68 line 21 is re-written as follows:

Referring to Figure [[12A]] 11 A, the cysteine-rich repeats of matriptase (aa 280-314, aa 315-351, aa 352-387, and aa 394-430) (SEQ ID NO: 15) are compared with the consensus sequences of the human LDL receptor (SEQ ID NO: 16); LDL receptor-related protein (LRP) (SEQ ID NO: 17); human perlecan (SEQ ID NO: 18); and rat GP-300 (SEQ ID NO: 19). The consensus sequences are boxed. In Figure [[12B]] 11 B, C1r/s type sequences of matriptase (aa 42-155 and aa 168-268) (SEQ ID NOS: 20 and 21) are compared with selected domains of human complement subcomponent Clr (aa 193-298) (SEQ ID NO: 22), C1s (aa 175-283) (SEQ ID NO: 23), Ra-reactive factor (RaRF) (aa 185-290) (SEQ ID NO: 24), and a calcium dependent serine protease (CSP) (aa 181-289) (SEQ ID NO: 25). The most consensus sequences are boxed.

The paragraph beginning on page 69 line 4 is re-written as follows:

The amino-terminal region of matriptase (aa 42-268) (SEQ ID NOS: 20 and 21) contains another two tandem segments with internal homology. These segments resemble

partial sequences, originally identified in complement subcomponents C1r (Leytus *et al.*, *Biochemistry* 25: 4855-63 (1986); and Journet *et al.*, *Biochem. J.* 240: 783-7 (1986)) and C1s (Mackinnon *et al.*, *Eur. J. Biochem.* 169: 547-53 (1987); and Tosi *et al.*, *Biochemistry* 26: 8516-24 (1987)). This C1r/s domain was also found in other serine proteases, including Ra-reactive factor, a C4/C2-activating component, enteropeptidase, an activator of trypsinogen (Matsushima *et al.*, (1994); Kitamoto *et al.*, (1994)), and a calcium-dependent serine protease that is able to degrade extracellular matrix. These C1r/s-containing serine proteases appear to be involved either in a protease activation cascade or in extracellular matrix degradation. In addition, there are at least six members of the astacin subfamily of zinc metalloprotease which were found to contain this C1r/s domain. These include bone morphogenetic protein-1 (Wozney *et al.*, *Science* 242: 1528-34 (1988)), and *Drosophila tolloid* gene, a dorsal-ventral patterning protein (Shimell *et al.*, Cell 67: 469-81 (1991)), quail 1, 25-dihydroxyvitamin D3-induced astacin like metallopeptidase that may play a role in the degradation of eggshell matrix, sea urchin blastula protease-10 (that could be involved in the differentiation of ectodermal lineages and subsequent patterning of the embryo), Xenopus embryonic protein UVS.2, a marker for developmental stage, and sea urchin VEB gene that is expressed in a spatially restricted pattern during the very early blastula stage of development. The majority of these C1r/s-containing, astacin metalloproteases appear to play a role in protein-protein interactions and embryonic development. The C1r/s domain has been also found in nonprotease proteins. These include neuropilin (A5 protein), a calcium-independent cell adhesion molecule that is developmentally-expressed in the nervous system and tumor necrosis factor-inducible protein TSG-6, a hyaluronate-binding protein that may be involved in cell-cell and cell-matrix interaction during inflammation and tumorigenesis.

The paragraph beginning on page 78 line 12 is re-written as follows:

To determine the position of the cleavage site for the generation of the two-chain form of matriptase, the 45- and 25-kDa components were each subjected to [[N-terininal]] N-terminal amino acid sequence analyses. The amino acid residues obtained from the 25-kDa B chain were VVGGTDADEGEWP (SEQ ID NO: 37). This sequence begins with the likely cleavage site within the activation motif in matriptase. When the 45-kDa A chain (including two major plus one minor spots) was sequenced, two overlapping sequences (SFVVTSVVAF

PTDSKTVQRT (SEQ ID NO: 38); TVQRTQDNSCSFGLHARGVE (SEQ ID NO: 39)) were obtained, and both matched sequences close to the amino terminus of matriptase. These two different amino-terminal sequences may be derived from the two major spots of matriptase A chain and suggest that the different migration rates of the two components result from their different amino termini.